Spring 5-31-2019

Deep morphological neural networks

Yucong Shen
New Jersey Institute of Technology

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

DEEP MORPHOLOGICAL NEURAL NETWORKS

by
Yucong Shen

Mathematical morphology is a theory and technique applied to collect features like geometric and topological structures in digital images. Determining suitable morphological operations and structuring elements for a give purpose is a cumbersome and time-consuming task. In this paper, morphological neural networks are proposed to address this problem. Serving as a non-linear feature extracting layers in deep learning frameworks, the efficiency of the proposed morphological layer is confirmed analytically and empirically. With a known target, a single-filter morphological layer learns the structuring element correctly, and an adaptive layer can automatically select appropriate morphological operations. For high level applications, the proposed morphological neural networks are tested on several classification datasets which are related to shape or geometric image features, and the experimental results have confirmed the tradeoff between high computational efficiency and high accuracy.

DEEP MORPHOLOGICAL NEURAL NETWORKS

by
Yucong Shen

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
In Partial Fulfillment of the Requirements for the Degree of
Master of Science in Computer Science

Department of Computer Science

May 2019

APPROVAL PAGE

DEEP MORPHOLOGICAL NEURAL NETWORKS

Yucong Shen

Dr. Frank Y. Shih, Thesis Advisor Professor of Computer Science, NJIT	Date
Dr. Zhi Wei, Committee Member Associate Professor of Computer Science, NJIT	Date
Dr. Hai Phan, Committee Member Assistant Professor of Informatics, NJIT	Date

BIOGRAPHICAL SKETCH

Author: Yucong Shen

Degree: Master of Science

Date: May 2019

Undergraduate and Graduate Education:

- Master of Science in Computer Science,
  New Jersey Institute of Technology, Newark, NJ, 2019

- Bachelor of Science in Mathematics and Applied Mathematics,
  Central China Normal University, Wuhan, China, 2017

Major: Computer Science

Presentations and Publications:

F.Y. Shih, Y. Shen and X. Zhong, "Development of deep learning framework for mathematical morphology," *Int. J. Pattern Recognit. and Artificial Intell.*, vol. 33, no. 6, pp. 1954024, June 2019.

Y. Shen, X. Zhong and F.Y. Shih, "Deep morphological neural networks," under submission to *IEEE trans Image Process.*.

DEDICATION

To Prof. Shih,

I not only learned the approach of doing research from you, but also the integrity and commitment.

To my family,

Thank you for your supports in all these years.

To Prof. Gaofeng Zheng & Prof. Xiao-Fei Zhang,

You set a great model to me when doing research. Thanks to you, I learned how to do research and the attitude towards acdemics.

To my dear motherland,

The rest of mine will be dedicated to you.

ACKNOWLEDGMENT

Appreciate Prof. Frank Y. Shih, my thesis advisor, for guiding me for almost two years in this fantastic area. Thanks for the committee member, Prof. Zhi Wei and Prof. Hai Phan. They gave me great help on doing research and how to make my thesis work better. Thanks Dr. Xin Zhong for giving me great help and many valuable suggestions in these two years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Mathematical morphology, which is based on set theory, can extract features based on shapes, regions, edges, skeleton, and convex hull [12]. The elementary operations in mathematical morphology are dilation and erosion, which are enlarging and shrinking the object respectively. Mathematical morphology has a wide range of applications in defect extraction [3], edge detection [19], and image segmentation [13]. In computer vision problems, deep learning has become increasingly popular in recent years. LeNet [7] was proposed for document recognition and digital recognition. Recently, the development of computer hardware brings the increased computational capacity, and CNN is becoming deeper, making CNN success on many applications of computer vision tasks, especially image recognition [5,17].

It is a time-consuming and cumbersome task to determine the proper morphological operations and the corresponding structuring elements. Shih *et al.* proposed MorphNet [16] to combine the advantages of mathematical morphology and deep learning to solve such problems, and also to provide a non-linear feature extractor for deep learning framework. The history of morphological neural network can be dated back to 1990s. Ritter *et al.* [9] proposed the morphological neural network based on image algebra [10]. It shows the first attempt in formulating useful morphological neural network. With respect to the linear feature extractor of the convolutional layer in CNN, MorphNet provides a morphological layer by the approximation of maximum and minimum, which simulate dilation and erosion operations respectively. With the help of morphological layer, we can determine the proper shape of structuring elements corresponding to the specific morphological operations, and also capture the non-linear

features of the image contents, especially the shapes.

Masci *et al.* [9] represented the dilation and erosion in deep learning framework using counter-harmonic mean. But they can only represent pseudo-dilation and pseudo-erosion due to the limitation of the formula. In [16], MorphNet can represent dilation and erosion accurately, and learn the binary structuring elements roughly, but failed in learning the non-flat structuring elements. Besides, it limits in learning the shape of corresponding structuring elements of dilation and erosion, and cannot determine the proper morphological operations applied on the original images.

In this paper, we propose the task of learning the structuring elements of two elementary operations of mathematical morphology, dilation and erosion. With the improvement of the MorphNet, we can learn the correct structuring elements by single-layer morphological neural network. Within dilation and erosion, we adopt a smooth sign function and a hyperbolic tangent function to determine the morphological operation by a single adaptive morphological layer. With the framework of morphological neural network consists of adaptive morphological layer, further applications of determining the morphological operations can be explored. What's more, because of the great property of mathematical morphology in extracting shapes features of image contents, we also propose a novel morphological layer based pipeline which captures the information of shapes.

Our key contribution can be summarized as follows. We present morphological layers for learning the correct binary morphological structuring elements and equivalent gray scale structuring elements. A morphological residual neural network architecture is developed for shape classification. We propose an adaptive morphological layer that can easily determine the proper morphological operations from a bunch of input and desired output images.

The remainder of this paper is organized as follows. Chapter 2 introduces the morphology layers, and the morphological residual neural network for shape classifications. Chapter 3 presents an adaptive morphological layer for determining the proper morphology operations from the original images and the desired result images. Chapter 4 shows the experimental results. Finally, conclusions are drawn in Chapter 5.

CHAPTER 2

DEEP MORPHOLOGICAL MORPHOLOGICAL NEURAL NETWORK

In this section, we illustrate the improvements of MorphNet and the approach to learn the corresponding structuring elements of morphological operations. We present an adaptive morphological neural network to provide a tool to learn the proper morphological operations from a bunch of original images and target images.

2.1 Previous Work

Masci *el al.* [9] represented the dilation and erosion in deep learning framework using counter-harmonic mean. For a grayscale image $f(x)$ and a kernel $\omega(x)$, the PConv layer performs as below:

$$PConv(f;\omega,P)(x) = \frac{(f^{P+1}*\omega)(x)}{(f^{P}*\omega)(x)} = (f *_P \omega)(x) \tag{2.1}$$

where "$*$" denotes the convolution operation, P is a scalar which controls the type of operation ($P < 0$ pseudo-erosion, $P > 0$ pseudo-dilation and $P = 0$ standard liner convolution). Since P cannot be infinity, this equation cannot represent the real erosion and dilation.

Shih *et al.* [16] represented the dilation and erosion using the soft maximum and soft minimum function. With the differential approximation of dilation and erosion. In dilation layer, the j-th pixel in the s-th feature map $z \in \mathbb{R}^n$, the size of the structuring elements (w.r.t. weights) is $n = a \times b$, then

$$z_j^s = \ln(\textstyle\sum_{i=1}^{n} e^{\omega_i x_i}) \tag{2.2}$$

where x_i is the i-th element of the masked window of the input images, and ω_i is the i-th element of the current weight.

Similarly, in the erosion layer, the j-th pixel in the s-th feature map $z \in \mathbb{R}^n$, the size of the structuring elements (w.r.t. weights) is $n = a \times b$, then

$$z_j^s = -\ln(\sum_{i=1}^{n} e^{-\omega_i x_i}), \tag{2.3}$$

where x_i is the i-th element of the masked window of the input images, and ω_i is the i-th element of the sliding window. Eqs. (2.2) and (2.3) are rough approximations of dilation and erosion, which can simulate the dilation and erosion more accurately. In MorphNet, approximating dilation and erosion is a much more straightforward and efficient way as compared to [9].

2.2 The Weakness of MorphNet

MorphNet also has a weakness when learning the correct structuring elements, although it represents the more accurate dilation and erosion theoretically. In [16], when single layer MorphNet learns the structuring elements from the input images and output images, there are missing points on the learned structuring elements with respect to the certain 3×3 and 5×5 structuring elements applied to original images. Fig. 2.1 shows the architecture of a single layer MorphNet when learning the structuring elements. Fig. 2.2 shows the structuring elements learned by the single dilation layer MorphNet. It shows that the single dilation layer MorphNet can learn part of the structuring elements as original ones, there are some biases between them.

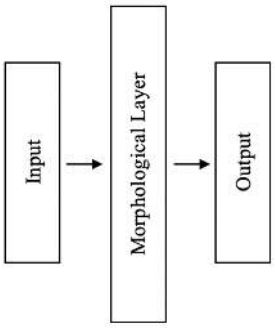


Figure 2.1 Architecture of the single layer MorphNet.

It can be observed from the representation of morphological layer of MorphNet, the only trainable parameter is the weight, the structuring element. The soft maximum does not round off the corner of when computing the maximum pixels of the sliding window, which results in biases from the original maximum pixels, and neither does soft minimum function. This causes the biases between the original structuring elements when creating the target images and the learned structuring elements.

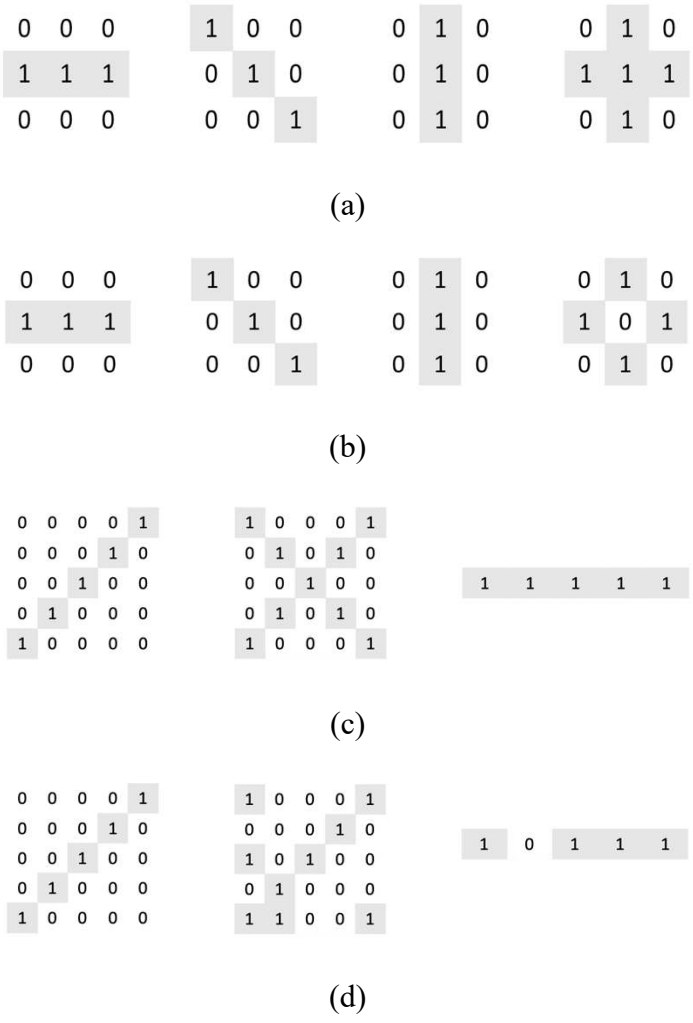


Figure 2.2 The original structuring elements applied on input images and the structuring elements learned by the single dilation layer MorphNet. (a) The horizontal, diagonal, vertical, and diamond 3×3 structuring elements applied to input images when creating target images, (b) the corresponding structuring elements learned by the single dilation layer MorphNet, (c) the original $45°$, crossing 5×5 structuring elements and horizontal line 1×5 structuring elements applied to the inputs images when creating target images, (d) the corresponding structuring elements learned by the single dilation layer MorphNet.

Definition 1 (Soft dilation): The j-th pixel of the result image $z \in \mathbb{R}^n$, is

$$z_j = \ln(\sum_{i=1}^{n} e^{\omega_i x_i}), \tag{2.4}$$

where ω is the structuring element, ω_i is the i-th element of the structuring element, the size of ω is $n = a \times b$, x_i is the i-th element of the masked window of the original image.

We call it *soft dilation*, and it can be denoted as $\omega \oplus x$, where the structuring element $\omega \in \mathbb{R}^n$, the original image $x \in \mathbb{R}^n$.

Definition 2 (Soft erosion): The j-th pixel of the result image $z \in \mathbb{R}^n$, is

$$z_j = -\ln(\sum_{i=1}^{n} e^{-\omega_i x_i}), \tag{2.5}$$

where ω is the structuring element, ω_i is the i-th element of the structuring element, the size of ω is $n = a \times b$, x_i is the i-th element of the masked window of the original image.

We call it *soft erosion*, and it can be denoted as $\omega \ominus x$, where the structuring element $\omega \in \mathbb{R}^n$, the original image $x \in \mathbb{R}^n$.

MorphNet follows *soft dilation* and *soft erosion*, we will show that the weakness of them in theory. For soft dilation, when learning the binary dilation, there should have

$$\ln(\sum_{i=1}^{n} e^{\omega_i x_i}) = \max(\omega_i x_1, \omega_2 x_2, \dots, \omega_n x_n), \tag{2.6}$$

indicates that

$$\ln(\sum_{i=1}^{n} e^{\omega_i x_i}) \geq x_i, \tag{2.7}$$

then we have

$$\sum_{i=1}^{n} e^{\omega_i x_i} \geq e^{x_i}. \tag{2.8}$$

Clearly, equation (2.8) is invalid. Therefore, we adopt a constant ζ, makes equation (2.8) shows as:

$$\sum_{i=1}^{n} e^{\omega_i x_i} \zeta \geq e^{x_i}, \tag{2.9}$$

then (2.9) is valid when $\zeta \geq \frac{e^{x_i}}{\sum_{i=1}^{n} e^{\omega_i x_i}}$.

We omit *soft erosion* in that it is similar to *soft dilation*.

2.3 The Improvement of MorphNet

Inspired by the convolutional neural network and equation (2.9), we introduce the bias to offset the bias caused by the soft maximum and soft minimum function. Therefore, in the dilation layer, the s-th feature map of the output z of dilation layer will be:

$$z^s = \omega \oplus x + b, \tag{2.10}$$

where $\omega \in \mathbb{R}^n$ is the weights, $x \in \mathbb{R}^n$ is the input of dilation layer, and $b \in \mathbb{R}^n$ is bias.

Similarly, the erosion layer can be expressed by:

$$z^s = \omega \ominus x + b, \tag{2.11}$$

where $\omega \in \mathbb{R}^n$ is the weights, $x \in \mathbb{R}^n$ is the input of dilation layer, and $b \in \mathbb{R}^n$ is bias.

After the improvement, it easy to get equation (2.12) follows the mathematics in section B:

$$(\sum_{i=1}^{n} e^{\omega_i x_i}) \cdot e^{b} \geq e^{x_i}. \tag{2.12}$$

Therefore, (2.12) will be valid if $b \geq \ln \frac{e^{x_i}}{\sum_{i=1}^{n} e^{\omega_i x_i}}$. Due to b is a trainable variable, so the dilation layer will be correct if $b \geq \ln \frac{e^{x_i}}{\sum_{i=1}^{n} e^{\omega_i x_i}}$ after the training when learning the binary dilation. Then we get the proof correctness of erosion layer when learning the binary erosion in the same way.

The gradient of such a layer is computed by back-propagation [6] with chain rule. The objective function can be denoted by $J(\omega, b; y, \hat{y})$, where ω is the weight, b is the bias, $\hat{y}$ is the output of the network, and y is the label of the network. Below is the gradient $\delta^{(l)}$ of the l-th layer of the network with respect to weight ω:

$$\delta^{(l)} = \frac{\partial J(\omega, b; y, \hat{y})}{\partial \omega^{(l)}}. \tag{2.13}$$

Assume that the learning rate is η, the weight ω of the l-th layer will be updated by:

$$\omega_i^{(l)} = \omega_i^{(l)} - \eta\delta^{(l)}, \tag{2.14}$$

the bias b will also be updated by back-propagation as:

$$b^{(l)} = b^{(l)} - \eta \frac{\partial J(\omega,b;y,\hat{y})}{\partial b^{(l)}}. \tag{2.15}$$

We name the neural networks that consists of morphological layers as Deep Morphological Neural Network (DMNN).

2.4 Learning the Morphological Operations by DMNN

We present the approach of learning binary and gray scale mathematical morphology operations and their corresponding structuring elements in this section.

2.4.1 Learning the Structuring Elements of Single Morphology Operation

We've proved the correctness of improved morphological layer in learning the binary morphology operations and their corresponding structuring elements. Here we showed the condition that the improved morphological layers can correctly learn the gray scale morphology operations and their corresponding structuring elements.

When learning the gray scale dilation, dilation layer, similar to (2.6), there should have

$$\ln(\sum_{i=1}^{n} e^{\omega_i x_i}) + b = \max(\omega_i + x_1, \dots, \omega_n + x_n). \tag{2.16}$$

From (2.16), we can easily get

$$e^b \cdot \sum_{i=1}^{n} e^{\omega_i x_i} \geq e^{x_i+\omega_i}, \tag{2.17}$$

then (2.17) is valid when $b \geq \ln \frac{e^{x_i+\omega_i}}{\sum_{i=1}^{n} e^{\omega_i x_i}}$. b is a trainable variable, so dilation layer will be correct if b maintains the condition after the training when learning the gray scale dilation. Similar proof can be applied to erosion layer for learning the gray scale erosion.

When learning the single binary and gray scale morphology operations, the architecture of the single layer morphological neural network still follows Fig. 2.1. The network minimizes the distance between the prediction of network and the target images.

2.4.2 Learning Multiple Morphology Operations

With the help of morphological layers, we can learn the multiple morphology operations by constructing multi-layer DMNN.

Assume the l-th layer of multi-layer DMNN is dilation layer, the s-th feature map of the output $z \in \mathbb{R}^n$ of current layer will be:

$$z_s^{(l)} = \omega \oplus z^{(l-1)} + b, \tag{2.18}$$

where $\omega \in \mathbb{R}^n$ is the weight of current layer, and $b \in \mathbb{R}^n$ is the bias, $z^{(l-1)} \in \mathbb{R}^n$ is the output of $(l-1)$-th layer.

If the l-th layer of multi-layer DMNN is erosion layer, the s-th feature map of the output $z \in \mathbb{R}^n$ of current layer will be:

$$z_s^{(l)} = \omega \ominus z^{(l-1)} + b, \tag{2.19}$$

where $\omega \in \mathbb{R}^n$ is the weight of current layer, and $b \in \mathbb{R}^n$ is the bias, $z^{(l-1)} \in \mathbb{R}^n$ is the output of $(l-1)$-th layer.

The inputs are the original images while the outputs are the predictions of network after multiple morphological layers. The target images are created by sequence of morphological operations. The deep morphological neural network will determine the proper corresponding structuring elements by learning from the input and target images, and will minimize the distance between outputs of the network and target images. After the network converges, the weights of each morphological layer will be the proper structuring elements for each morphological operation. The deep

morphological neural networks which consist of few stacked morphological layers can learn the morphological operations pipeline and determine the proper structuring element for each step.

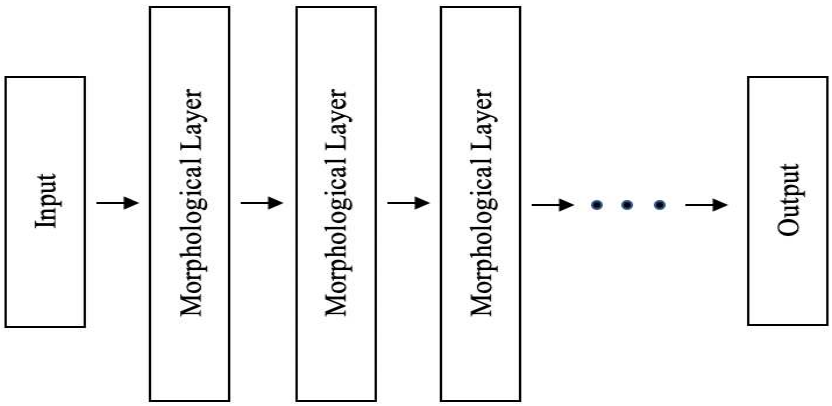


Figure 2.3 Architecture of the multi-layer deep morphological neural network.

The gradient of multi-layer DMNN is also computed by back-propagation with chain rule. The objective function can be denoted by $J(\omega, b; y, \hat{y})$, where ω is the weight, b is the bias, $\hat{y}$ is the output of the network, and y is the label of the network. The gradient $\delta^{(l)}$ of the l-th layer with respect to weight ω:

$$\delta^{(l)} = \frac{\partial J(\omega,b;y,\hat{y})}{\partial \omega^{(l)}} = \frac{\partial J(\omega,b;y,\hat{y})}{\partial z^{(l)}} \frac{\partial}{\partial \omega} \sigma'\left(z^{(l)}\right), \tag{2.20}$$

where $\sigma(\cdot)$ is the activation function.

Assume that the learning rate is η, the weight ω of the l-th layer will be updated by:

$$\omega_i^{(l)} = \omega_i^{(l)} - \eta \delta^{(l)}. \tag{2.21}$$

2.5 Morphological Residual Neural Network

Mathematical morphology always comes with shapes and structures [4, 18] in the applications. In pattern recognition, mathematical morphology is being used for preprocessing and feature extraction.

In morphological residual model, applying opening on the original image with circle structuring elements, then the edges of the shapes will become smoothly. Then after the subtraction of result image from original images, we can get the residuals. After morphological residual, we obtain the residuals of the geometric shapes. Fig. 2.4 shows how the morphological residual model extracts the residuals from geometric shapes. With the help of morphological residual, it is easy to recognize the shapes by counting the number of residuals.

Figure 2.4 The morphological residual model. Applying opening on the original image with circle structuring elements, then subtraction of result image from original image can obtain the morphological residuals.

2.5.1 The Architecture of Morphological Residual Neural Network

Followed by the morphological residual model, we construct morphological residual neural network for shape classification. Fig. 2.5 shows the architecture of the morphological residual neural network. The input of the neural network are batches of images, erosion layer and dilation layer are applying opening to the input images. After the subtraction layer, the neural network finishes the preprocessing progress and delivers the residuals to classifier. Before the classifier, there are two fully-connected layers for flattening the matrix to column vector and also data compression. At the end of the network, a softmax classifier is classifying the images and producing the outputs.

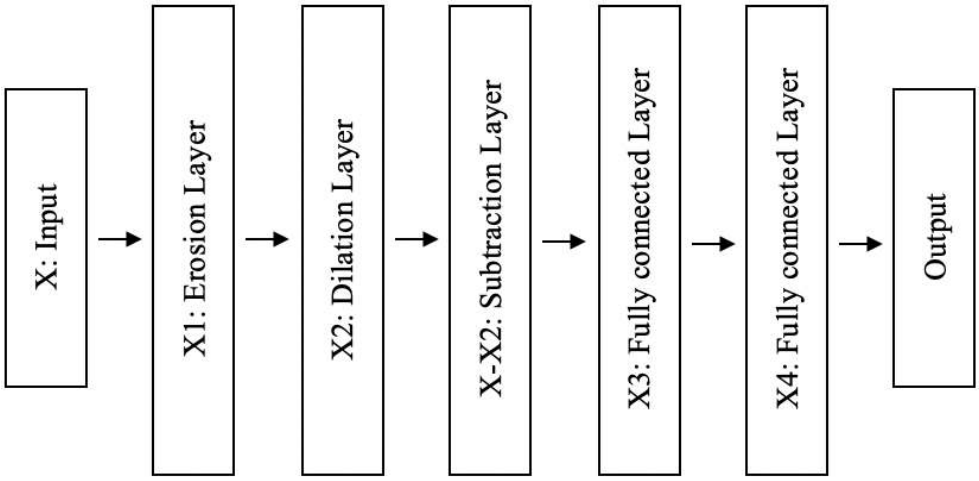


Figure 2.5 The architecture of morphological residual neural network.

The configuration of the morphological residual neural network is followed by the modern convolutional neural network. In the erosion layer and dilation layer, we adopt 3×3 filter size to reduce the parameters. In the first three weights layer, the channel is 1 for grayscale images, and 3 for RGB images. Table 2.1 shows the configuration of the morphological residual neural network in detail.

Table 2.1 The Configuration of Morphological Residual Neural Network

	Input
1	Erosion $3 \times 3 \times 1$
2	Dilation $3 \times 3 \times 1$
3	Subtraction 1
4	FC-1024
5	FC-512
6	Soft-max

2.5.2 The Gradient of Morphological Residual Neural Network

We also trained the morphological residual neural network by back-propagation. The weights of dilation layer, erosion layer and fully connected layers are updated by (2.20) and (2.21). In the subtraction layer, the weights will not be updated, it just transmits the gradient from fourth layer to second layer.

Assume the gradient of fourth layer is $\delta^{(4)}$, the gradient of subtraction layer will be:

$$\delta^{(3)} = \delta^{(4)}. \tag{2.22}$$

CHAPTER 3

ADAPTIVE MORPHOLOGICAL LAYER

In the applications of mathematical morphology, deciding the proper operation is also a tough and time-consuming task. Especially there are various of morphological operations, such as dilation, erosion, opening, closing, etc. It is not difficult to make the decision on choosing dilation or erosion due to dilation enlarges the object in the image when erosion shrinks the object in the image. Yet, it is a time-consuming task when makes the decisions on large scale images dataset due to various features of the images and the needs.

We can observe from the expression soft maximum and soft minimum functions that they are extracting maximum and minimum pixels in same way except soft minimum is the opposite of soft maximum. Therefore, the j-th pixel on the output $z \in \mathbb{R}^n$ of the dilation and erosion layer (we can name it as adaptive morphological layer) can be represented by:

$$z_j = sign(\mathrm{x}) \cdot \ln(\sum_{i=1}^{n} e^{sign(x) \cdot \omega_i x_i}) + b, \tag{3.1}$$

where a is an extra trainable variable aside from ω_i and b. If $sign(x)$ is $+1$, the operation of current layer would be dilation; if $sign(x)$ is -1, the operation of current layer would be erosion; otherwise, the operation of current layer neither will be dilation nor erosion. However, the sign function is not a continuous function and not differential so it cannot be introduced to the neural network. Then the smooth sign function can be adopted to replace the sign function. Note that there are various functions smooth in the interval $[-1,1]$, such as soft sign function, hyperbolic tangent function, etc. Equation (3.2) and (3.3) show the soft sign function and hyperbolic tangent function.

$$f(x) = \frac{x}{1+|x|} \tag{3.2}$$

$$g(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3.3}$$

Therefore, we introduce hyperbolic tangent function and soft sign function to the adaptive morphological layer by replacing sign function with them. Then the j-th pixel on the output $z \in \mathbb{R}^n$ of the adaptive morphological layer in two ways:

$$z_j = \frac{a}{1+|a|} \cdot \ln\left(\sum_{i=1}^{n} e^{\frac{a}{1+|a|} \cdot \omega_i x_i}\right) + b, \tag{3.4}$$

or

$$z_j = \frac{e^a - e^{-a}}{e^a + e^{-a}} \cdot \ln\left(\sum_{i=1}^{n} e^{\frac{e^a - e^{-a}}{e^a + e^{-a}} \cdot \omega_i x_i}\right) + b, \tag{3.5}$$

where is a trainable variable, and $a \in \mathbb{R}$. ω_i is also the i-th pixel in the sliding window, b is the bias.

In the comparison of the soft sign function and hyperbolic tangent function, Fig. 3.1 shows the figure of soft sign function and tanh function, it can be observed that value of both functions lie on the interval $[-1,1]$. Hyperbolic tangent function reaches -1 and $+1$ ahead of soft sign function in that the value of soft sign function is around -0.8 when tanh function reaches -1, similarly, the value of soft sign function lies on around 0.8 when tanh function almost reaches $+1$. Clearly, the gradient of the soft sign is always smaller than hyperbolic tangent function from the figure. Therefore, hyperbolic tangent function increases faster than soft sign function. If not considering the computing speed, the hyperbolic tangent function will have better performance than the soft sign function theoretically.

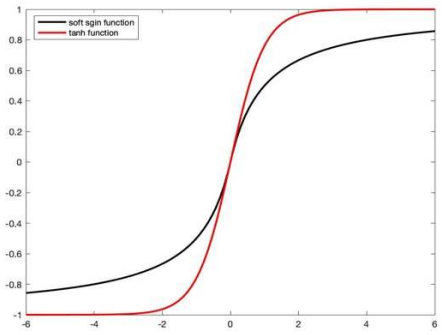


Figure 3.1 The soft sign function and hyperbolic tangent function.

Having the adaptive morphological layer, we can determine the correct single morphology operation by a single layer neural network, which consist of one adaptive morphological layer. The input of the single layer adaptive morphological neural network are the original images, the target images are dilated or eroded images. After feeding in batches of images data, the network minimizes the distance between network outputs and target images. If the value of soft sign function or hyperbolic tangent function approaches $+1$, the neural network will predict that the target images are dilated images; if the value of soft sign function or hyperbolic tangent function approaches -1, the neural network will predict that the target images are eroded images; otherwise, the neural network will predict that the target images are neither dilated image nor eroded images.

The gradient of the adaptive morphological neural network will also be updated by back-propagation with the chain rule. The weight is being updated by gradient descent, which is a typical optimization algorithm for neural network. Assume that the objective function of such neural network is $J(\omega, b, a; y, \hat{y})$, where ω is the weight, b is the bias, a is a trainable variable for indicating the morphological operations, $\hat{y}$ is the output of the network, and y is the label of the network. The gradient $\delta^{(l)}$ of the l-th layer with respect to weight a is:

$$\delta^{(l)} = \frac{\partial J(\omega,b,a;y,\hat{y})}{\partial a^{(l)}} = \frac{\partial J(\omega,b,a;y,\hat{y})}{\partial z^{(l)}} \frac{\partial z^{(l)}}{a^{(l)}} = \frac{\partial J(\omega,b,a;y,\hat{y})}{\partial z^{(l)}} \varphi'(a), \tag{3.6}$$

where $\varphi(\cdot)$ is the soft sign or hyperbolic tangent function.

Assume that the learning rate is η, the weight a of the l-th layer will be updated by:

$$a^{(l)} = a^{(l)} - \eta\delta^{(l)}. \tag{3.7}$$

CHAPTER 4

EXPERIMENTAL RESULTS

Our implementation is done by the keras, the experiments are based on 4-GPU system, which equipped with four NVIDIA Titan X GPUs. We present our experimental results on MNIST, self-created geometric shapes dataset, GTSRB (German Traffic Sign Recognition Benchmark) dataset [15].

MNIST dataset is a database consist of 70,000 examples of handwritten digits 0~9. It has 60,000 training images, and 10,000 testing images. They are all 28×28 gray scale images in 10 classes.

Self-created geometric shapes dataset is created by Python PIL library due to the limited resources of public geometric shape database. The images in this database are all 64×64 grayscale images. There are 5 classes: ellipse, line, rectangle, triangle, and five-edge polygon. The white shape object is randomly drawn on a black background, their size, position, and orientation are randomly initialized. In the training set, each class has 20,000 images, 100,000 images in total. In the testing set, each class has 5,000 images, 20,000 images in total. Fig. 4.1 shows examples from this database.

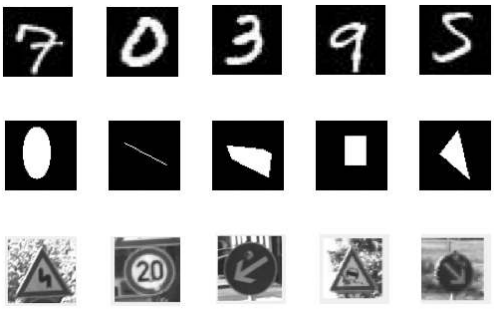


Figure 4.1 The sample image from three dataset we adopt in experiments. The first row shows images from MNIST. The second row are the images from self-created geometric shapes. The third row are the images from GTSRB dataset.

GTSRB (German Traffic Sign Recognition Benchmark) [15] is a single-image, multi-class classification problem, there are 42 classes in total. The images contain one traffic sign each and each real-world traffic sign only occurs once. We resized all the images into 31×35, and select 31,367 images for training, 7,842 images for testing. During the preprocessing, we converted all the images to grayscale images.

4.1 Results on Deep Morphological Neural Network

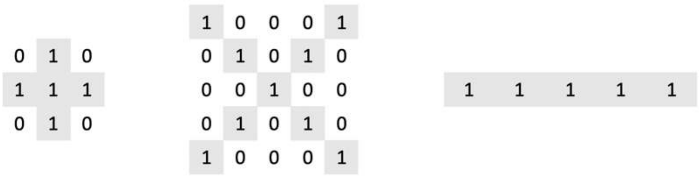
We present the experimental results on MNIST in this part. In our experiment, we applied 10,000 images for training progress.

4.1.1 Learning the Binary Structuring Elements

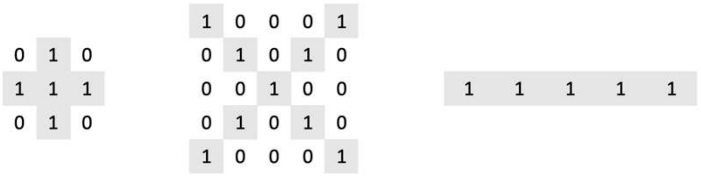
After applying the improvement of MorphNet, we can successfully correct the error in Fig. 2.2.

When learning a single binary structuring element, we construct single layer morphological neural network shown in Fig. 1, adopt MSE (Mean squared error) to measure the distance between the target images and predictions of the neural network. The target images are created by applying dilation or erosion on the original input images. We were minimizing the distance between predictions and target images by mini-batch SGD [8] with a batch size of 64, the learning rate is $\eta = 7.50$.

Fig. 4.2 shows the results after improvement, it is easy to see that Fig. 4.2 (a) and (b) are all the same.



(a)



(b)

Figure 4.2 The original structuring elements applied on input images and the structuring elements learned by the single dilation layer MorphNet after the improvement. (a) The diamond 3×3 structuring element, crossing 5×5 structuring element, horizontal line 1×5 structuring element applied to input images when created target images; (b) the corresponding structuring elements learned by single dilation layer morphological neural network after improvement.

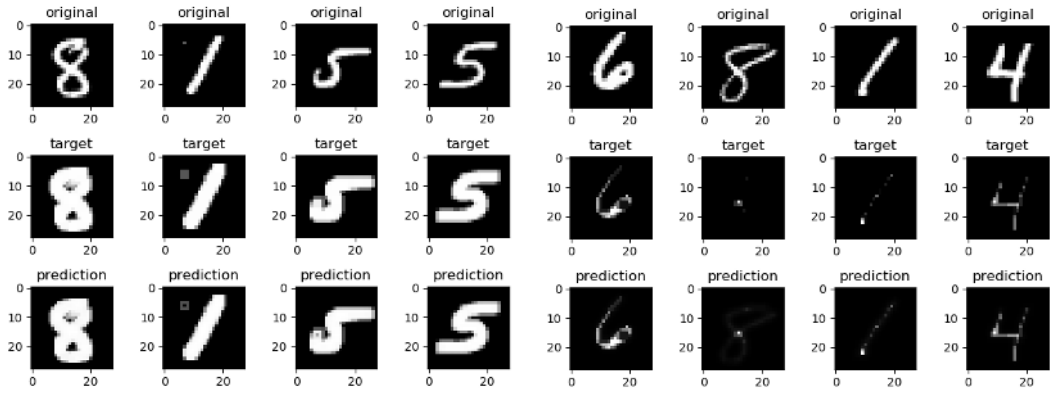
4.1.2 Learning the Gray Scale Structuring Elements

We showed the result of learning the binary structuring elements by improved morphological neural network. But the grayscale morphology differs from binary morphological operations [14], and we've proved that morphological layer can simulate the grayscale morphology in theory. Therefore, we will show the effectiveness of morphological neural network on grayscale morphology by experiments in this part.

Similar to the procedure of learning the binary structuring elements, the target images are created by applying dilation or erosion on the original input images, the distance between predictions of neural network and target images is measured by MSE (Mean squared error). We were minimizing the distance between predictions and target images by mini-batch SGD with a batch size of 64, the learning rate is $\eta = 10.00$ for learning eroded images, and $\eta = 7.50$ for learning dilated images.

After 20 epochs, the MSE loss was being minimized to be around 5.19×10^{-4} when training dilation. Fig. 4.3 (a) shows the result for learning dilation by single dilation layer morphological neural network. The target images and the output of the network is visually equal by human eyes. When the single erosion layer morphological neural network minimizes the MSE loss to be around 5.84×10^{-4}. Fig. 4.3 (b) shows

the result for single erosion layer morphological neural network. The single erosion layer morphological neural network can also learn the same output of the network as the target images.



(a) (b)

Figure 4.3 The result for gray scale structuring elements on MNIST dataset. The first row shows the original images, the second row shows the target images, and the third row shows the output of the network after training 20 epochs. (a) shows the result of learning dilation; (b) shows the result of learning erosion.

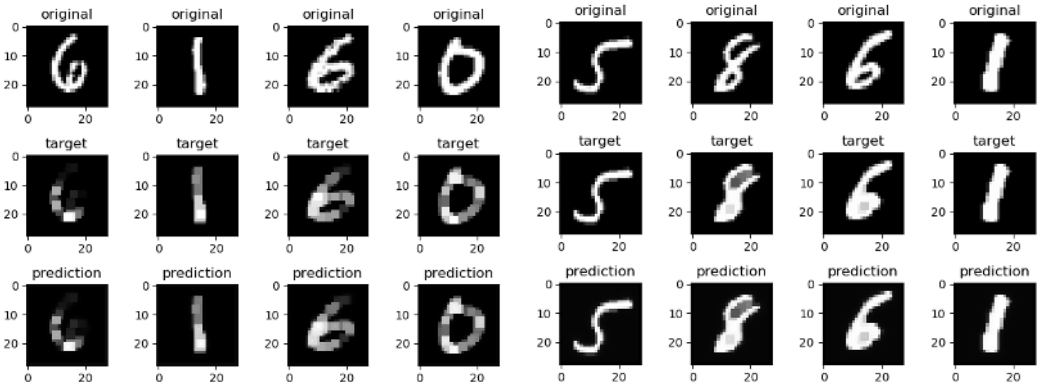
4.1.3 Learning the Multiple Morphological Operations by DMNN

In mathematical morphology, opening and closing are also important morphological operations. Assume that dilation is denoted by $A \cdot B$, where A is the original image and B is the structuring element, and erosion is denoted by $A \circ B$, where A is the original image and B is the structuring element. The opening will be denoted by $(A \circ B) \cdot C$, where A is the original image and B and C is the structuring element. The closing will be denoted by $(A \cdot B) \circ C$, where A is the original image and B and C is the structuring element. Therefore, we construct two-layer DMNNs to learn opening and closing operations.

When learning opening, we random initialize a 3×3 structuring element to create the target images. Then we construct a two-layer DMNN with an erosion layer, and a dilation layer after the erosion layer. When learning closing, we also random initialize a 3×3 structuring element to create the target images, then construct a two-

layer DMNN with a dilation layer, and an erosion layer after dilation layer. We learn opening and closing operations by these two DMNNs.

For training, we also adopted mini-batch SGD algorithm, the batch size is set to be 64, and the learning rate is $\eta = 10.0$. The loss will converge in 10 epochs when learning opening and closing. Fig. 4.4 shows the experimental results for learning opening and closing.



(a) (b)

Figure 4.4 The result for learning the opening and closing operations by DMNN. The first row shows the original images, the second row shows the target images, and the third row shows the output of the network after training 20 epochs. (a) shows the result of learning opening, (b) shows the result of learning closing.

From Fig. 4.4 (a) and (b), it is easy to see that the target images and predictions of DMNN is visually identical.

4.1.4 Morphological Residual Neural Network for Classification

We present our result of classification on MNIST, self-created geometric shapes, and GTSRB dataset.

For training, we used mini-batch algorithm, the batch size is 64 and the learning rate is $\eta = 0.0001$. We follow the architecture shown in Fig. 2.5, and the configuration in Table 3.1. The morphological residual can converge in 10 epochs when training on self-created geometric shape dataset, and converges in 70 epochs when training on

GTSRB dataset. The testing accuracy of the morphological residual is 98.89% on self-created geometric shape dataset, and 95.35% on GTSRB, and 98.93% on MNIST dataset. We added a dropout layer after the second fully-connected layer due to the overfitting problem when training on the GTSRB, the testing accuracy increased to 96.49%. Table 4.1 shows the configurations of morphological residual neural networks when training on three datasets, a indicates the number of filters applied in each layer.

Table 4.1 Comparison of Morphological Residual in Three Datasets

	MNIST	Self-created Geometric Shapes	GTSRB
Erosion layer	$3 \times 3 \times a$	$3 \times 3 \times a$	$3 \times 3 \times a$
Dilation layer	$3 \times 3 \times a$	$3 \times 3 \times a$	$3 \times 3 \times a$
Subtraction layer	$28 \times 28 \times a$	$64 \times 64 \times a$	$64 \times 64 \times a$
Fully-connected layer	120	1024	1024
Fully-connected layer	84	512	512
Output	10	5	43

Morphological residual has great classification rate on self-create dataset and real images dataset. We modify LeNet [15], and name it as Modified LeNet (MLeNet). Table 4.2 shows the configuration of MLeNet. In MLeNet, we add one more convolutional layer to extract more features, decrease the size of the filters from 5×5 to 3×3 to save parameters.

Table 4.2 Configuration of MLeNet

	Input
1	Convolutional layer $3 \times 3 \times 16$
2	Max pooling 2×2
3	Convolutional layer $3 \times 3 \times 32$
4	Max pooling 2×2
5	Convolutional layer $3 \times 3 \times 64$
6	Max pooling 2×2
7	Fully-connected 2048×1
8	Fully-connected 1024×1
9	Softmax

Table 4.3 Comparison of Morphological Residual with State-of-Art Convolutional Neural Network

Classifier	Dataset	Testing accuracy	Number of parameters
MCDNN [2]	MNIST	99.77%	2,682,470
Morphological residual	MNIST	98.93%	104,181
MLeNet	Self-created geometric shapes	99.50%	10,493,795
Morphological residual	Self-created geometric shapes	98.89%	4,721,175
MLeNet	GTSRB (Grayscale)	97.94%	4,202,339
Morphological residual	GTSRB (Grayscale)	96.49%	1,594,903

When $a = 1$, table 4.3 shows the comparisons with the state-of-the-art convolutional neural network on our current result. Although morphological residual loses on the testing accuracy compared to state-of-the-art convolutional neural network, morphological residual has much fewer parameters. We significantly cut off the number of parameters of the deep neural network and provide a tradeoff between the number of parameters and the testing accuracy. Especially in the feature extraction layers (the weights layers except for fully connected layers, such as convolutional layer and morphological layer), morphological residual has only 20 parameters in total, we show the comparison of the number of parameters in feature extraction layer of morphological residual with state-of-art convolutional neural network in table 4.4. From table 4.3 and 4.4, we can observe that morphological residual uses much fewer parameters in feature extraction, but does not lose too much accuracy compared to convolutional neural networks. Morphological residual has great a tradeoff between the efficiency of extracting features from image contents and testing accuracy.

We conclude that morphological residual is efficient in extracting features, and saves parameters of trainable weights of the neural net.

Table 4.4 Comparison of Number of Parameters in Feature Extraction Layer of Morphological Residual with State-of-Art Convolutional Neural Network

Model	Number of parameters in feature extraction layers
Morphological residual	20
MLeNet	2,912
MCDNN	739,900

Moreover, we show the advantages of morphological layers in shape classification task. We construct a CNN that has same configuration as morphological residual neural network, and compare its performance with morphological residual neural network on classification.

Table 4.5 shows the configuration of the CNN that we construct in the comparison with morphological residual neural network, we name it as residual CNN. b indicates the number of filters in each layer.

Table 4.5 Configuration of Residual CNN

	Input
1	Convolutional layer $3 \times 3 \times b$
2	Convolutional layer $3 \times 3 \times b$
3	Subtraction layer $3 \times 3 \times b$
4	Fully-connected $2048 \times b$
5	Fully-connected $1024\times b$
6	Softmax

Table 4.6 shows the comparison of residual CNN and morphological residual neural network on the classification tasks.

Table 4.6 Comparison of Morphological Residual and Residual CNN

	Morphological residual ($a = 1$)	Residual CNN ($b = 1$)	Morphological residual ($a = 16$)	Residual CNN ($b = 16$)
MNIST	98.93%	97.14%	97.78%	98.18%

Self-created geometric shapes	98.89%	98.25%	98.90%	98.91%
GTSRB	96.49%	90.60%	97.48%	93.39%

In table 4.6, when $a = 1$ and $b = 1$, morphological residual neural network has better testing accuracy on all three datasets than residual CNN; when a$= 16$ and $b = 16$, morphological residual has better testing accuracy on GTSRB dataset. Therefore, morphological layer performs better than convolutional layers if both neural networks have same structure. Especially on GTSRB dataset, morphological layer significantly improves the testing accuracy. It indicates that morphological layer has advantages on classifying shapes, and can gain shape information more efficient.

4.2 Results on Detecting Morphological Operations by Adaptive Morphological Neural Network

In this section, we random selected 10,000 images from the MNIST dataset for training. We applied dilation or erosion on the original images to get the target images (unknown to the neural network), the target images can be considered as desired result images in industrial applications.

We were using mini-batch SGD to optimize the network, the batch size is 64, and the learning rate is $\eta = 10.0$. We construct a morphological neural network consists of one adaptive morphological layer. We also measure the distance between the predictions and the target images by MSE loss.

After 20 epochs, the single adaptive morphological neural network converges, the MSE loss between the target images and the prediction decreases to around 3×10^{-4}. We test both (3.4) and (3.5) with the same configuration of the network,

optimization method, epochs, loss function and learning rate. When evaluating the results, if the value of the smooth sign function larger than 0.8 but not larger than 1.0, we would round it off to 1.0, if the value of the smooth sign function smaller than -0.8 but not smaller than -1.0, we would round it off to -1.0. Due to the properties of two smooth sign function we adopt, the value of smooth sign function will not exceed the interval $[-1,1]$.

We train the single adaptive morphological layer neural network 100 times each on two smooth sign function and dilated target images, eroded target images, it achieved 100% detection accuracy on detecting dilation and erosion both by soft sign function and hyperbolic tangent function.

CHAPTER 6

CONCLUSIONS

We have presented the framework of deep morphological neural network. After the improvement, the morphological layers can learn the correct binary structuring elements and equivalent non-flat structuring elements. We provide the architecture of morphological residual neural network for shape classification. Morphological residual neural network achieves a great tradeoff between model accuracy and number of parameters, and significantly decreases the model parameters. We also show the advantages of morphological layer in extracting shape features of objects in images. The adaptive morphological layer provides a tool to determine the proper morphology operations from original images and desired result images, the adaptive morphological neural network can automatically learn single morphology operation by a single adaptive morphological layer. Deep morphological neural network provides a non-linear feature extraction layer for deep learning frame work, and also solutions to cumbersome image morphology industrial applications.

REFERENCES

[1] R. Coliban, M. Ivanovici and N. Richard, "Improved probabilistic pseudo-morphology for noise reduction in colour images," in *IET Image Processing*, vol. 10, no. 7, pp. 505-514, 7 2016.

[2] Ciregan, U. Meier and J. Schmidhuber, "Multi-column deep neural networks for image classification," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, Rhode Island, 2012, pp.3642-3649.

[3] K. Chen, Z. Zhang, Y. Chao, M. Dai and J. Shi, "Defects extraction for QFN based on mathematical morphology and modified region growing," *2015 IEEE International Conference on Mechatronics and Automation (ICMA)*, Beijing, China, 2015, pp. 2426-2430.

[4] R. M. Haralick, S. R. Sternberg and X. Zhuang, "Image Analysis Using Mathematical Morphology," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 4, pp. 532-550, July 1987.

[5] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," *IEEE conference on Computer Vision and Pattern Recognition (CVPR),* 2016, pp. 770-778.

[6] Y. LeCun, B.E. Boser, J.S. Denker, D. Henderson, R.E. Howard, W.E. Hubbard and L.D. Jackel, "Handwritten digit recognition with a back-propagation network", In *Advances in neural information processing systems*, 1990, pp.396-404.

[7] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

[8] M. Li, T. Zhang, Y. Chen and A.J. Smola, "Efficient mini-batch training for stochastic optimization," *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*, New York, NY, USA, pp. 661-670.

[9] J. Masci, J. Angulo and J. Schmidhuber, "A learning framework for morphological operators using counter-harmonic mean," *Proc. Internatinal Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing*, Springer, Berlin, Heidelberg, 2013, pp. 329-340.

[10] G. X. Ritter and P. Sussner, "An introduction to morphological neural networks," *Proceedings of 13th International Conference on Pattern Recognition*, Vienna, Austria, 1996, pp. 709-717 vol.4.

[11] G.X. Ritter and J.N. Wilson, "*Handbook of Computer Vision Algorithms in Image Algebra*," CRC press, 1996.

[12] F. Y. Shih, "Image processing and mathematical morphology: fundamentals and applications," CRC press, 2009.

REFERENCES

[13] H. Shih and E. Liu, "Automatic Reference Color Selection for Adaptive Mathematical Morphology and Application in Image Segmentation," in *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4665-4676, Oct. 2016.

[14] F.Y. Shih and O. R. Mitchell, “Threshold decomposition of gray-scale morphology into binary morphology,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 1, pp. 31-42, Jan. 1989.

[15] J. Stallkamp, M. Schlipsing, J. Salmen and C. Igel, "The German Traffic Sign Recognition Benchmark: A multi-class classification competition," *The 2011 International Joint Conference on Neural Networks*, San Jose, CA, 2011, pp. 1453-1460.

[16] F.Y. Shih, Y. Shen and X. Zhong, “Development of deep learning framework for mathematical morphology,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 6, pp. 1954024, June 2019.

[17] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *International Conference on Learning Representation*, San diego, CA, USA, 2015.

[18] F. Zana and J.C. Klein, “Segmentation of vessel-like patterns using mathematical morphology and curve evaluation,” *IEEE transactions on image processing*, 2001, 10(7), pp.1010-1019.

[19] W. Zhang, D. Shi and X. Yang, "An improved edge detection algorithm based on mathematical morphology and directional wavelet transform," *2015 8th International Congress on Image and Signal Processing (CISP)*, Shenyang, China, 2015, pp. 335-339.